Webinar Transcript
Evidence for ESSA: An Introduction to Study Design

Slide No. 1:

Welcome to today's webinar: "Evidence for ESSA – An Introduction to Study Design." This webinar is brought to you by District 180 in the Office of Continuous Improvement and Support at the Kentucky Department of Education.

Slide No. 2:

Here are our objectives for today:

By the end of this webinar, participants will be able to…

- describe the basic principles of study design;
- define key words related to different study designs; and
- determine the significance of study findings.

Slide No. 3:

The agenda for this webinar is on the screen. We will begin with an introduction to study design and a discussion regarding this webinar's alignment to existing statutes, regulation, and guidance. Then, we will discuss the three broad categories of study design mentioned in the Every Student Succeeds Act (ESSA) – experimental study design, quasi-experimental study design and correlational study design. Finally, we will review the principles of statistical significance.

Slide No. 4:

Before we get started, let's take a moment to consider why an understanding of study design is helpful. In 2015, the U.S. Congress reauthorized the Elementary and Secondary Education Act through a bill known as the Every Student Succeeds Act (ESSA for short). One of the requirements of ESSA is that school improvement initiatives be rooted in "evidence-based activities, strategies, or interventions." Before you can begin to apply ESSA's evidence provisions, you must first develop a general understanding of study design and how different study designs impact the investigation of the effectiveness of an intervention.

Slide No. 5:

Study designs provide a framework for the development and implementation of a study. A study is a detailed investigation and analysis of a subject or situation. The study design framework guides researchers as they collect and analyze data to test solutions and solve problems.

Study designs are important because they provide a common language that helps researchers and policymakers interpret and discuss research findings in a meaningful way. By understanding the basic principles of study design, a layperson can better engage in a conversation about the findings of a study and its potential impact.

There are many study designs available to researchers, and each study design has a different purpose. Researchers select study designs that are most appropriate for their outcome, study methodology, study topic or research capacity. Some study designs are more rigorous than others, and it is important to understand how this variety may impact the outcome of a study.

This webinar will focus on three broad study designs specifically mentioned in the Code of Federal Regulations. They are the experimental, quasi-experimental and correlational study designs.

Slide No. 6:

Effort has been made to ensure that the definitions and key concepts presented in this webinar align to multiple educational resources. Those references include the Code of Federal Regulations, “Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments” and the “What Works Clearinghouse Standards Handbook” (Version 4.0). All three resources have been hyperlinked in the PowerPoint for your convenience. The table on the screen is from the non-regulatory guidance and the three study designs featured in this webinar are listed in the first row of the table.

Slide No. 7:

We will begin our discussion with the most rigorous study design – the experimental study design.

Slide No. 8:

An experimental study design, also called a randomized control trial or RCT, is a study design in which two randomly assigned groups of study participants are compared to determine if an intervention was successful. One study group, the intervention group, receives the intervention. The other group, the control group, does not receive the intervention. By comparing the two groups, researchers are able to measure the effect of an intervention.

Slide No. 9:

Let’s break that definition down into some key points. First, experimental studies utilize random assignment when placing study participants into intervention or control groups. This means that individual participants are sorted into groups completely by random chance, such as a lottery or coin flip. This is the hallmark of an experimental study design. By randomly assigning participants to groups, the researcher is able to eliminate common variables that may influence the outcome of a study. In education, these variables may include things like socioeconomic status, the experience of the teacher or the individual backgrounds of the students.

Experimental studies examine the impacts of an intervention by comparing the work of an intervention group to the work of a control group. This comparison allows researchers to evaluate the intervention without other outside influences.

For example, if a study evaluates the impact of paired reading on student comprehension but only examines the progress of students in one class who all received the intervention, then there is no way to know for sure that paired reading actually influenced student comprehension. Other factors may have had an impact, such as the time of day students were taught or the quantity of reading materials available to them.

The standards used vet study designs in education are set by the What Works Clearinghouse, or WWC. The WWC evidence standards allow researchers to treat study participants as individuals or clusters in experimental studies. This means that researchers can randomly assign individuals to groups, or they can randomly assign clusters of individuals (such as classrooms) to groups.

The combination of random assignment and the comparison of two groups make experimental studies the most rigorous study design that we will be discussing in this webinar. Consequently, experimental studies are linked to the highest level of evidence in ESSA.

Slide No. 10:

When evaluating a study that used an experimental study design, there are two things to consider. The first is a compromised random assignment. Random assignment is compromised when students are assigned to groups based on conditions or switch groups after random assignment. For example, a study's random assignment would become compromised if a principal switched low-performing students from the control group to the intervention group once the study has begun. In this instance, the principal's decision to move students may have impacted the outcome of the study.

Another common concern for experimental studies is the rate of attrition. Attrition is the progressive loss of data or subjects during a research study. Attrition creates bias that may impact the outcome of a study. Both overall attrition and differential attrition – that is, attrition that occurs within the intervention and control groups – are important for interpreting the results of a study. The WWC sets the standard for acceptable rates of attrition using the Attrition and Potential Bias table on the screen.

Bias created by a compromised random assignment or high levels of sample attrition can be addressed through statistical adjustments by the study authors or may be treated as a quasi-experimental study.

Slide No. 11:

Next, we will explore the characteristics of the quasi-experimental study design.

Slide No. 12:

A quasi-experimental study is a study design in which two previously assigned groups of study participants are compared to determine if an intervention was successful. One study group, the intervention group, receives the intervention. The other group, the control group, does not receive the intervention. By comparing the two groups, researchers are able to measure the effect of an intervention.

Slide No. 13:

The quasi-experimental study design is similar to the experimental study design in many ways. Both designs compare an intervention group to a control group and both designs allow study participants to be treated as individuals or as clusters. The primary difference is that the quasi-experimental study design uses previously assigned groups – in other words – the groups are not randomly assigned.

For example, a quasi-experimental study may compare the results of fifth-grade students broken into two classes. This is not a random assignment because schools use a variety of tools and strategies to place students into classrooms. This process creates a bias. Quasi-experimental studies must use statistical controls to address this bias.

Slide No. 14:

One of the most common forms of bias in quasi-experimental study design is related to the selection of the groups. Groups that are previously assigned typically have similarities that influenced their assignment. To correct for this, researchers must establish baseline equivalence between the two groups. This statistical correction shows that the intervention group and the control group had characteristics that were similar enough at the start of the study. Common statistical adjustments to look for include regression adjustments, analysis of covariance and difference-in-difference adjustments.

Quasi-experimental studies are also heavily prone to confounding factors. A confounding factor is a characteristic that is aligned to one group but not the other. This is very common in studies that assign classrooms to different conditions, because schools may group students by characteristics – such as putting lower-performing students with a more experienced teacher or assigning English learners (ELs) with a teacher who holds an extra credential. This alignment creates a bias that influences the outcome of the study.

The lack of random assignment and high rate of confounding factors are what makes quasi-experimental study designs less rigorous than experimental study designs. This is why studies that use quasi-experimental designs are considered Level 2 evidence by ESSA.

Slide No. 15:

The third study design we will be discussing today is the correlational study design.

Slide No. 16:

A correlational study design relies on observational data (collected by the researcher in a natural environment without interference), archival data (publically available data reported by local and state education agencies) or survey data (collected by the researcher through anonymous surveying) to draw a statistical, or correlational, conclusion.

Slide No. 17:

The correlational study design is one that is rooted in the use of observational, archival or survey data. Correlational studies are very popular among education researchers due to the large amount of publically available data and the relative ease of surveying large populations of teachers, parents or students.

In a correlational study, the researcher attempts to measure the relationship between two variables. For example, a correlational study may seek to determine if there is a relationship between student reading fluency and the number of behavior referrals posted by a school. A strong correlation between low levels of reading fluency and high numbers of behavior referrals may indicate that we can improve one variable by improving the other.

It is important to note that correlation does not equal causation. Rather, it demonstrates that a relationship exists between two variables and that further investigation is necessary.

Correlational studies are commonly used to study the impact of whole school, district and/or state intervention efforts, efforts to improve teacher quality and the effect of socio-emotional interventions on student outcomes.

Slide No. 18:

When evaluating a correlational study, here are a few things to consider:

The correlational study design is a very detached and objective methodology. It takes the data for what it is and looks for relationships. The researcher should pay careful attention to ensure that variables are not influenced or manipulated during the collection process.

Just as with the other study designs mentioned in this webinar, bias is a potential issue that must be addressed. In correlational research, two common types of bias are confirmation bias and sampling bias. Confirmation bias occurs when a researcher designs a study in such a way as to confirm a hypothesis. This is common in surveying, where questions may be worded in such a way as to lead study participants to a certain answer. Sampling bias occurs when a researcher selects certain types of data in hopes of finding certain answers. This is an easy mistake to make when using archival data, where the researcher must manually select the data sources and data points to be extracted.

Finally, correlational studies are math-driven studies. While all of the study designs mentioned in this webinar may fall victim to inaccurate calculations, correlational studies are more prone to these types of mistakes. While you can generally trust that the calculations presented in peer-reviewed research have been validated, it is always good to look at findings with a careful eye.

Slide No. 19:

Now that we have explored three common research designs, let's examine some other important concepts related to the significance of study findings.

Slide No. 20:

First, let's consider the sampling process. The analytic sample is the sample on which an analysis is based. The sampling process includes the population, setting and sample size used in a given study. Generally speaking, the larger and more diverse a study is, the more reliable the results are likely to be.

"Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments" suggests a large sample size of 350 or more students or 50 or more groups of 10 or more students as the ideal sample size for educational research. It is also desirable for an analytic sample to encompass more than one site wherein site means a local education agency, locality or state.

You should also look at the setting and population that participated in a study. ESSA's evidence provisions indicate that the more closely aligned the setting and population of an analytic sample is to your school, the stronger the evidence.

Slide No. 21:

Once you have examined the sampling process used in a study, take a look at the measures used. The performance measure is a quantitative indicator, statistic or metric used to gauge program or project performance. In other words, it is how the researchers know whether or not an intervention worked.

The performance measure should be aligned to a relevant outcome. A relevant outcome is the student outcome – or the ultimate outcome if not related to students – the proposed process, product, strategy or practice is designed to improve; consistent with the specific goals of a program.

Using these two measures, the researchers will draw a causal inference – or a conclusion that an activity or intervention was likely to have affected an outcome.

Slide No. 22:

The results of an experimental, quasi-experimental or correlational study will include statistical analysis. Here we have presented some common statistical terms that you may see in educational research.

Reliability and validity are two terms that refer to the quality of the instrument used to collect data. Reliability refers to the dependability or consistency of an instrument, while validity refers to the quality or soundness of an instrument. Most researchers will seek to use a collection instrument that has high levels of reliability and validity. Standardized tests are considered to be both reliable and valid.

The results of education research are often presented using three common statistical descriptors – standard deviation, p-value and effect size. Standard deviation represents the variability of a measure across the observations of as sample. A low standard deviation means that the data points are close together, while a high standard deviation means that the data points are more spread apart.

The p-value is a statistical reporting measure used to describe outcome significance. The relationship between two variables is said to be significant if p is less than 0.05. An outcome with a p-value of less than 0.05 is said to "reject the null hypothesis," or the hypothesis that no relationship exists between two variables.

Finally, the effect size presents a standardized measures of the magnitude of a difference. If the p-value tells us whether or not a relationship is significant, the effect size tells us how much. Effect size is reported using the effect size index, or d value, sometimes called Cohen's *d*. It is commonly accepted that an effect size is small if d equals 0.20, medium if d equals 0.50, and large if d equals 0.80.

Effect size and standard deviation are often reported together, with both using a bell curve graph.

Slide No. 23:

For more detailed information about study design and how it impacts ESSA's evidence provisions, we encourage you to complete the WWC Group Design Standards Online Training provided by the What Works Clearinghouse. Links and other references are provided on the next slide for this webinar.

Slide No. 24:

Thank you for viewing this webinar. During the webinar, we provided you with an introduction to study design; a discussion on the alignment of study design principles to ESSA's evidence provisions; the key characteristics and potential pitfalls of experimental, quasi-experimental and correlational study designs and a brief overview of the key concepts used in measuring significance. Resources that were consulted during this presentation are hyperlinked for your convenience.

Slide No. 25:

If you have questions regarding evidence-based interventions or study design, please contact the District 180 branch in the Office of Continuous Improvement and Support at (502) 564-2116.